

MICRODATA CONFIDENTIALITY REFERENCES W. E. Winkler 2004.March.04

- Abowd, J. M. and Woodcock, S. D. (2002), "Disclosure Limitation in Longitudinal Linked Data," in (P. Doyle et al, eds.) *Confidentiality, Disclosure, and Data Access*, North Holland: Amsterdam.
- Abowd, J. M. and Woodcock, S. D. (2004), "Multiply-Imputing Confidential Characteristics and File Links in Longitudinal Linked Data, in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York
- Adams, N. R. and Wortmann, J. C., (1989), "Security-control Methods for Statistical Databases, A Comparative Study," *ACM Computing Surveys*, **21**, 515-556.
- Agrawal, D. and Aggarwal, C. C. (2001), "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," Association of Computing Machinery, Proceedings of ACM SIGPODS, 247-255.
- Agrawal, R. and Srikant, R. (2000), Privacy Preserving Data Mining, Proceedings of the ACM SIGMOD, 439-450.
- Agrawal, R., Kiernan, J., Srikant, R. and Xu, Y. (2002), "Hippocratic Databases," *Very Large Databases '02*.
- Bacher, J., Bender, S. and Brand, R. (2001), "Re-identifying Register Data by Survey Data: An Empirical Study," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Bacher, J., Brand, R. and Bender, S. (2002) Re-identifying Register Data by Survey Data using Cluster Analysis: An Empirical Study, *International Journal of Uncertainty, Fuzziness, Knowledge-Based Systems*, 10 (5) 589-608
- Bethlehem, J. A., Keller, W. J. and Pannekoek, J., (1990) "Disclosure Control of Microdata," *Journal of the American Statistical Association*, **85**, 38-45.
- Blien, U., Wirth, U. and Muller, M. (1992), "Disclosure Risk for Microdata Stemming from Official Statistics," *Statistica Neerlandica*, **46**, 69-82.
- Brand, R. (2002), "Microdata Protection Through Noise Addition," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.
- Castro, J. (2004), "Computational Experience with Minimum-Distance Controlled Perturbation Methods," in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York
- Dalenius, T. and Reiss, S.P. (1982), "Data-swapping: A Technique for Disclosure Control," *Journal of Statistical Planning and Inference*, **6**, 73-85.
- Dandekar, R. A. (2004), Maximum Utility Minimum Information Loss Table Server Design of Statistical Disclosure Control of Tabular Data, in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York.
- Dandekar, R. A., Domingo-Ferrer, J. and Sebe, F. (2002), "LHS-Based Hybrid Microdata vs Rank Swapping and Microaggregation for Numeric Microdata Protection," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.
- Dandekar, R., Cohen, M. and Kirkendal, N. (2002), "Sensitive Microdata Protection Using Latin Hypercube Sampling Technique," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.
- Defays, D. and Anwar, M. N. (1998), "Masking Microdata Using Micro-aggregation," *Journal of Official Statistics*, **14**, 449-461.
- Defays, D. and Nanopolis, P. (1993), "Panels of Enterprises and Confidentiality: the Small Aggregates Method," in Proceedings of the 1992 Symposium on Design and Analysis of Longitudinal Surveys, 195-204.
- De Waal, A. G. and Willenborg, L.C.R.J. (1995), "Global Recodings and Local Suppressions in Microdata Sets," *Proceedings of Statistics Canada Symposium 95*, 121-132.
- De Waal, A. G. and Willenborg, L.C.R.J. (1996), "A View of Statistical Disclosure Control for Microdata," *Survey Methodology*, **22**, 95-103.
- Domingo-Ferrer, J. (2001), "On the Complexity of Microaggregation," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Domingo-Ferrer, J. (ed.) (2002) *Inference Control in Statistical Databases*, Springer: New York
- Domingo-Ferrer, J. and Mateo-Sanz, J. M. (2001), "An Empirical Comparison of SDC Methods for Continuous Microdata in Terms of Information Loss And Re-Identification Risk," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Domingo-Ferrer, J. and Mateo-Sanz, J. M. (2002), "Practical Data-Oriented Microaggregation for Statistical Disclosure Control," *IEEE Transactions on Knowledge and Data Engineering*, **14** (1), 189-201.
- Domingo-Ferrer, J., Mateo-Sanz, J., Oganian, A. and Torres, A. (2002), "On the Security of Microaggregation with Individual Ranking: Analytic Attacks," *International Journal of*

- Uncertainty, Fuzziness, and Knowledge-Based Systems*, 477-492.
- Domingo-Ferrer, J., Sebé, F., and Castellà-Roca, J. (2004), "On the Security of Noise Addition for Privacy In Statistical Databases, in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York.
- Domingo-Ferrer, J. and Torra, V. (2001) A Quantitative Comparison of Disclosure Control Methods for Microdata in (P. Doyle, J. Lane, J. Theeuwes, and L. Zayatz, eds.) *Confidentiality, Disclosure Control and Data Access: Theory and Practical Applications*, North Holland, 111-134.
- Domingo-Ferrer, J. and Torra, V. (2003), "Statistical Data Protection in Statistical Microdata Protection Via Advanced Record Linkage," *Statistics and Computing*, **13** (4), 343-354.
- Duncan, G. T., Keller-McNulty, S. A., and Stokes, S. L. (2001), "Disclosure Risk vs. Data Utility: The R-U Confidentiality Map," Los Alamos National Laboratory Technical Report LA-UR-01-6428.
- Elamir, E. A. H. (2004), "Analysis of Re-identification Risk based on Log-Linear Model," in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York
- Elliott, M. A., Manning, A. M., and Ford, R. W. (2002), "A Computational Algorithm for Handling the Special Uniques Problem, *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, **10** (5), 493-510.
- Elliot, M.J., Skinner, C. A., and Dale, A. (1998), "Special Uniques, Random Uniques, and Sticky Populations: Some Counterintuitive Effects of Geographic Detail in Disclosure Risk," *Research in Official Statistics*, **1**, 53-68.
- Evfimievski, A., Srikant, R., Agrawal, R. and Gehrke, J. (2002), "Privacy Preserving Mining of Association Rules," *Association of Computing Machinery, Special Interest Group on Knowledge Discovery and Datamining '02*.
- Fellegi, I. P., and Sunter, A. B. (1969), "A Theory for Record Linkage," *Journal of the American Statistical Association*, **64**, 1183-1210.
- Fellegi, I. P. (1997), "Record Linkage and Public Policy - A Dynamic Evolution," *Proceedings of the Record Linkage Workshop 1997*, Washington, DC: National Academy Press, 3-12.
- Fienberg, S. E. (1997), "Confidentiality and Disclosure Limitation Methodology: Challenges for National Statistics and Statistical Research, commissioned by Committee on National Statistics of the National Academy of Sciences.
- Fienberg, S. E., Makov, E. U. and Sanil, A. P., (1997), "A Bayesian Approach to Data Disclosure: Optimal Intruder Behavior for Continuous Data," *Journal of Official Statistics*, **14**, 75-89.
- Fienberg, S. E., Makov, E. U. and Steel, R. J. (1998), "Disclosure Limitation using Perturbation and Related Methods for Categorical Data," *Journal of Official Statistics*, **14**, 485-502.
- Frakes, W. and Baeza-Yates, R. (1992), "Information Retrieval - Data Structures and Algorithms," Prentice-Hall: Upper Saddle River, N.J.
- Franconi, L., Capobianchi, A., Polletini, S. and Seri, G. (2001), "Experiences in Model-Based Disclosure Protection," presented at the UNECE Workshop on Statistical Data Confidentiality, Skopje, Macedonia, May 2001.
- Fuller, W. A. (1993), "Masking Procedures for Microdata Disclosure Limitation," *Journal of Official Statistics*, **9**, 383-406.
- Gopal, R., Goes, P. and Garfinkel, R. (1998) "Confidentiality Via Camouflage: The CVC Approach to Database Query Management," in *Statistical Data Protection '98*, Eurostat, Brussels, Belgium, 1-8. (also (2002) *Operations Research*, **50** (3))
- Gill, L. (1999), "OX-LINK: The Oxford Medical Record Linkage System," in *Record Linkage Techniques 1997*, Washington, DC: National Academy Press, 15-33.
- Gomatam, S. V. and Karr, A. (2003), "On Data Swapping of Categorical Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, to appear.
- Gouweleeuw, J.M., Kooiman, P., Willenborg, L. C. R. J. and De Wolf, P.-P. (1998), "Post Randomisation For Statistical Disclosure Control: Theory and Implementation," *Journal of Official Statistics*, **14**, 463-478.
- Grim, J., Bocek, P. and Pudil, P. (2001), "Safe Dissemination of census Results by Means of Interactive Probabilistic Models," *Proceedings of 2001 NITS and ETK*, Eurostat: Luxembourg, 849-856.
- Hwang, J. T. (1986), "Multiplicative Error-in-Variables Models with Applications to Recent Data Released by the U.S. Department of Energy," *Journal of the American Statistical Association*, **81** (395), 680 - 688.
- Iyengar, V. (2002), "Transforming Data to Satisfy Privacy Constraints," Association of Computing Machinery, Special Interest Group on Knowledge Discovery and Datamining '02.

- Kennickell, A. B. (1999), "Multiple Imputation and Disclosure Control: The Case of the 1995 Survey of Consumer Finances," in *Record Linkage Techniques 1997*, Washington, DC: National Academy Press, 248-267 (available at <http://www.fcsn.gov>).
- Kargupta, H., Datta, S., Wang, Q, and Ravikumar, K. (2003) "Random Data Perturbation Techniques and Privacy Preserving Data Mining," Expanded version of best paper awarded paper from the IEEE International Conference on Data Mining, November, 2003, Orlando, FL.
- Kim, J. J. (1986), "A Method for Limiting Disclosure in Microdata Based on Random Noise and Transformation," American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 303-308.
- Kim, J. J. (1990), "Subdomain Estimation for the Masked Data," American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 456-461.
- Kim, J. J. and Winkler, W. E. (1995), "Masking Microdata Files," American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 114-119.
- Kim, J. J. and Winkler, W. E. (2001), "Multiplicative Noise for Masking Continuous Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, cd-rom.
- Lambert, D. (1993), "Measures of Disclosure Risk and Harm," *Journal of Official Statistics*, **9**, 313-331.
- Lawrence, C., Zhou, J.L. and Tits, A. L. (1997), "User's Guide for CFSZP Version 2.5: A C Code for Solving (Large Scale) Constrained Nonlinear Inequality Constraints," Unpublished, Electrical Engineering Dept. and Institute for Systems Research, University of Maryland.
- Liew, C. K., Choi, U. J. and Liew, C. J. (1991), "A Data Distortion by Probability Distribution," *ACM Transactions on Database Systems*, **10**, 395-411.
- Little, R. J. A. (1993), "Statistical Analysis of Masked Data," *Journal of Official Statistics*, **9**, 407-426.
- Little, R. J. A. and Liu, F. (2002), "Selective Multiple Imputation of Keys for Statistical Disclosure Control in Microdata," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Little, R. J. A. and Liu, F. (2003), "Comparison of SMiKe with Data-Swapping and PRAM for Statistical Disclosure Control of Simulated Microdata," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, to appear.
- Malin, B. Sweeney, L., and Newton, E. (2003), "Trail Re-identification: Learning Who You are from Where You have Been, Workshop on Privacy in Data, Carnegie-Mellon University, March 2003.
- Mera, R. (1998), "Matrix Masking Methods That Preserve Moments," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 445-450.
- Moore, R. (1995), "Controlled Data Swapping Techniques For Masking Public Use Data Sets," U.S. Bureau of the Census, Statistical Research Division Report rr96/04, (available at <http://www.census.gov/srd/www/byyear.html>).
- Müller, W., Blien, U. and Wirth, H. (1995), "Identification Risks of Micro Data," *Evidence from Experimental Studies. Sociological Methods and Research*, **24**, 131-157.
- Muralidhar, K., Batrah, D. and Kirs, P.J. (1995), "Accessibility, Security, and Accuracy in Statistical Databases : The Case for the Multiplicative Fixed Data Perturbation Approach," *Management Science*, **41**(9), 1549-1584
- Muralidhar, K., Parsa, R. and Sarathy, R. (1999), "A General Additive Data Perturbation Method for Database Security," *Management Science*, **45** (10), 1399-1415.
- Muralidhar, K. and Sarathy, R. (1999) "Security of Random Data Perturbation Methods," *ACM Transactions on Database Systems*, **24** (4), 487-493.
- Muralidhar, K. and Sarathy, R. (2003), "A Theoretical Basis for Perturbation Methods," *Statistics and Computing*, **13** (4), 329-335.
- Muralidhar, K., Sarathy, R. and R. Parsa, R. (2001) "An Improved Security Requirement for Data Perturbation with Implications for E-Commerce," *Decision Sciences*, **32** (4), 683-698.
- Paas, G. (1988), "Disclosure Risk and Disclosure Avoidance for Microdata," *Journal of Business and Economic Statistics*, **6**, 487-500.
- Palley, M. A. and Simonoff, J. S. (1987), "The Use of Regression Methodology for the Compromise of Confidential Information in Statistical Databases," *ACM Transactions on Database Systems*, **12** (4), 593-608.
- Polletini, S. (2003), "Maximum Entropy Simulation for Microdata Protection," *Statistics and Computing*, **13** (4), 307-320.
- Polletini, S., Franconi, L., and Stander, J. (2002), "Model Based Disclosure Protection," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*: New York.
- Polletini, S. and Stander, J. (2004), "A Bayesian Hierarchical Model Approach to Risk Estimation in

- Statistical Disclosure Limitation,” in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York
- Raghunathan, T. E. (2003), Evaluation of Inferences from Multiple Synthetic Data Sets Created Using Semiparametric Approach, Panel on Confidential Data Access for Research Purposes, Committee On National Statistics, October 2003.
- Raghunathan, T. E., Lepkowski, J. M., Hoewyk, J. V. and P. Sollenberger (1998), “A Multivariate Technique for Multiply Imputing Missing Values Using a Series of Regression Models,” Survey Research Center, University of Michigan.
- Raghunathan, T.E., Reiter, J. P. and Rubin, D.R. (2003), “Multiple Imputation for Statistical Disclosure Limitation,” *Journal of Official Statistics*, **19**, 1-16.
- Raghunathan, T.E. and Rubin, D.R. (2000), “Multiple Imputation for Disclosure Limitation” technical report
- Reiss, J.P. (1984), “Practical Data Swapping: The First Steps,” *ACM Transactions on Database Systems*, **9**, 20-37.
- Reiter, J.P. (2002), “Satisfying Disclosure Restrictions with Synthetic Data Sets,” *Journal of Official Statistics*, **18**, 531-543.
- Reiter, J.P. (2003a), “Inference for Partially Synthetic, Public Use Data Sets,” *Survey Methodology*, to appear.
- Reiter, J.P. (2003b), “Releasing Multiply Imputed, Synthetic Public-Use Microdata: An Illustration and Empirical Study,” *Journal of the Royal Statistical Society, A*.
- Reiter, J.P. (2003c), Estimating Probabilities of Identification for Microdata, Panel on Confidential Data Access for Research Purposes, Committee On National Statistics, October 2003.
- Roque, G. M. (2000), “Masking Microdata Files with Mixtures of Multivariate Normal Distributions,” Ph.D.Dissertation, Department of Statistics, University of California at Riverside.
- Rubin, D. B. (1993), “Satisfying Confidentiality Constraints through the Use of Synthetic Multiply-imputed Microdata,” *Journal of Official Statistics*, **91**, 461-468.
- Samarati, P. (2001), “Protecting Respondents’ Identity in Microdata Release,” *IEEE Transactions on Knowledge and Data Engineering*, **13** (6), 1010-1027.
- Samarati, P. and Sweeney, L. (1998), “Protecting Privacy when Disclosing Information: k-anonymity and Its Enforcement through Generalization and Cell Suppression,” Technical Report, SRI International.
- Sarathy, R. and Muralidhar, K. (2002), “The Security of Confidential Numerical Data in Databases,” *Information Systems Research*, **48** (12), 1613-1627.
- Sarathy, R., Muralidhar, K., and Parsa, R. (2002), Perturbing Non-Normal Attributes: The Copula Approach, *Management Science*, **48** (12), 1613-1627
- Scheuren, F. and Winkler, W. E. (1993), “Recursive Merging and Analysis of Administration Lists,” American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 123-128 (presently available on <http://www.amstat.org> in the Section on Government Statistics).
- Scheuren, F. and Winkler, W. E. (1997), “Regression Analysis of Data Files that are Computer Matched – Part II,” *Survey Methodology*, 157-165).
- Schlörer, J. (1981), “Security of Statistical Databases: Multidimensional Transformation,” *ACM Transactions on Database Systems*, **6**, 91-112.
- Skinner, C. J. and Elliot, M. A. (2001), “A Measure of Disclosure Risk for Microdata,” *Journal of the Royal Statistical Society*, **64** (4), 855-867.
- Skinner, C. J. and Holmes, D. J. Estimating the Re-identification Risk per Record in Microdata, *Journal of Official Statistics*, **14** (1998) 361-372
- Stander, J. and Franconi, L. (2001), “A Model-Based Disclosure Limitation Method for Business Microdata,” presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Sullivan, G. and Fuller, W. A. (1989), “The Use of Measurement Error to Avoid Disclosure,” American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 802-807.
- Sullivan, G. and Fuller, W. A. (1990), “Construction of Masking Error for Categorical Variables,” American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 435-439.
- Sweeney, L. (1999), “Computational Disclosure Control for Medical Microdata: The Datafly System” in *Record Linkage Techniques 1997*, Washington, DC: National Academy Press, 442-453.
- Sweeney, L. (2002), “Achieving k-Anonymity Privacy Protection Using Generalization and Suppression,” *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 571-588.
- Takemura, A. (2002), “Local Recoding and Record Swapping by Maximum Weight Matching for Disclosure Control of Microdata Sets,” *Journal of Official Statistics*, **18** (2), 275-289.
- Tendick, P. and N. Matloff (1994), “A Modified Random Perturbation Method for Database

- Security," *ACM Transactions on Database Systems*, **19**, 47-63.
- Thibaudeau, Y. (2004), "An Algorithm for Computing Full Rank Minimal Sufficient Statistics with Application to Confidentiality Protection, *UNECE Statistical Journal*, to appear.
- Thibaudeau, Y. and Winkler, W.E. (2002), "Bayesian Networks Representations, Generalized Imputation, and Synthetic Microdata Satisfying Analytic Restraints," Statistical Research Division report RR 2002/09 at <http://www.census.gov/srd/www/byyear.html>.
- Torra, V. and Miyamoto, S. (2004), "Evaluating Fuzzy Clustering Algorithms for Microdata Protection," in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York
- Trottini, M. and Fienberg, S. E. (2002), "Modelling User Uncertainty for Disclosure Risk and Data Utility," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 511-528.
- Van Den Hout, A. and Van Der Heijden, P. G. M. (2002), "Randomized Response, Statistical Disclosure Control, and Misclassification: A Review, *International Statistical Review*, **70**(2), 269-288.
- Van Gewerden, L., Wessels, A., and Hundepol, A. (1997), "Mu-Argus Users Manual, Version 2," Statistics Netherlands, Document TM-1/D.
- Willenborg, L. and De Waal, T. (1996), *Statistical Disclosure Control in Practice*, Vol. 111, Lecture Notes in Statistics, Springer-Verlag, New York.
- Willenborg, L. and De Waal, T. (2000), *Elements of Statistical Disclosure Control*, Vol. 155, Lecture Notes in Statistics, Springer-Verlag, New York.
- Winglee, M., Valliant, R., Clark, J., Lim, Y., Weber, M. and Strudler, M. (2002), "Assessing Disclosure Protection for the SOI Public Use File," American Statistical Association, *Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Winkler, W. E. (1994), "Advanced Methods for Record Linkage, American Statistical Association, *Proceedings of the Section on Survey Research Methods*, pp. 467-472.
- Winkler, W. E. (1995), "Matching and Record Linkage," in B. G. Cox (ed.) *Business Survey Methods*, New York: J. Wiley, 355-384.
- Winkler, W. E. (1997), "Views on the Production and Use of Confidential Microdata," Statistical Research Division report RR 97/01 at <http://www.census.gov/srd/www/byyear.html>.
- Winkler, W. E. (1998), "Re-identification Methods for Evaluating the Confidentiality of Analytically Valid Microdata," *Research in Official Statistics*, **1**, 87-104.
- Winkler, W. E. (2002a), "Using Simulated Annealing for k-anonymity," Statistical Research Division report RR 2002/07 at <http://www.census.gov/srd/www/byyear.html>.
- Winkler, W. E. (2002b), "Single Ranking Micro-aggregation and Re-identification," Statistical Research Division report RR 2002/08 at <http://www.census.gov/srd/www/byyear.html>.
- Winkler, W. E. (2004a), "The State of Masking and Re-identification Methods for Public-Use Microdata: Overview and Research Problems, in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York.
- Winkler, W. E. (2004b), "Re-identification Methods for Masked Microdata, in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York.
- Yancey, W.E., Winkler, W.E., and Creecy, R. H. (2002) "Disclosure Risk Assessment in Perturbative Microdata Protection," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.